

# Supplementary Appendix

## Estimated impact of screening on gonorrhoea epidemiology in the United States: insights from a mathematical model

### 1 Model overview

We constructed a deterministic compartmental model of gonorrhoea transmission in the United States population. The population was divided into compartments representing the following states: susceptible (S), infectious and symptomatic (Y), and infectious and asymptomatic (Z). The model also included a never sexually active (A) compartment. The model was stratified by sex (male and female), sexual activity group (low and high), age category (15-24 y and 25-39 y), and subpopulation. The four subpopulations were: heterosexual non-Hispanic blacks, heterosexual Hispanics, heterosexual ‘white and others’ (non-Hispanic non-black population), and gay, bisexual, and other men who have sex with men (MSM). The model included individuals aged 15-39 years of age. In 2015, 89% of all reported gonorrhoea cases in the US occurred in this age group [1]. The choice of racial/ethnic groups was based on observed rates, which are elevated in non-Hispanic blacks and Hispanics relative to non-Hispanic whites [1], as well as the fact that these two population groups represent the largest racial and ethnic minority groups in the United States.

For simplicity, we assumed a constant population size ( $N$ ) and a closed population; that is, no sexual partnerships occurred with individuals outside of the modeled population. The proportion of the population in each racial/ethnic group ( $pop_{ij}$ ) was based on 2014 U.S. Census estimates [2]. We assumed an equal number of males and females in the total population. The proportion of males in the MSM group was based on a meta-analysis of US population-based surveys [3]. With the allocation of a proportion of males from each race/ethnic group to the MSM population, there were more females than males in each of the heterosexual subpopulations. For each sex, the size of the subpopulation was calculated as:  $N * pop_{ij}$ .

New individuals entered the 15-24 y age group in the susceptible state, with a proportion allocated to the never sexually active group, where they were not at risk of infection. As individuals aged into the older age category, a proportion of never sexually active individuals transitioned out of this compartment and became susceptible to gonorrhoea infection. Transitions out of the ‘never sexually active’ compartment as individuals aged were unidirectional; that is, once a person had their sexual debut they could not return to the ‘never sexually active’ compartment. We included this compartment to account for the fact

that in the younger age group in particular, the ever sexually active population is significantly smaller than the total population [4].

Susceptible individuals could be infected following a sexual partnership with an infectious individual and transitioned to symptomatic or asymptomatic infection. We included a time-varying relative risk of transmission in MSM to account for the possibility that risk behaviours in MSM may be increasing with the widespread availability of highly active antiretroviral therapy for treatment and prevention of HIV [5, 6]. The probability of having a symptomatic infection was assumed to be higher in males than females [7, 8]. Since oropharyngeal and rectal infections are more likely to be asymptomatic than urethral infections and are identified much more frequently in MSM [7, 9], we assumed that the probability of presenting with a symptomatic infection was lower in MSM than heterosexual males.

## 2 Model equations and parameters

For an individual of a given subpopulation ( $i$ ), sex ( $j$ ), sexual activity group ( $k$ ), and age group ( $l$ ), the model is described by the following system of differential equations, where  $N_{ijkl}$  represents the total sexually active population in a given group. Parameter definitions are provided in Table 1 of this appendix. Details on calculation of the sexual mixing matrix and force of infection are provided in the subsequent sections.

$$\begin{aligned}\frac{dS_{ijk1}}{dt} &= -\lambda_{ijk1}S_{ijk1} + \gamma_{ij}Y_{ijk1} + \delta_{ij}Z_{ijk1} + \varphi_{ijk1}Z_{ijk1} \\ &\quad + p_{s_{ijk1}}\mu_2(N_{ijk2} + A_{ijk2}) - \mu_1S_{ijk1}\end{aligned}$$

$$\frac{dY_{ijk1}}{dt} = \sigma_{ij}\lambda_{ijk1}S_{ijk1} - \gamma_{ij}Y_{ijk1} - \mu_1Y_{ijk1}$$

$$\frac{dZ_{ijk1}}{dt} = (1 - \sigma_{ij})\lambda_{ijk1}S_{ijk1} - \delta_{ij}Z_{ijk1} - \varphi_{ijk1}Z_{ijk1} - \mu_1Z_{ijk1}$$

$$\frac{dA_{ijk1}}{dt} = (1 - p_{s_{ijk1}})\mu_2(N_{ijk2} + A_{ijk2}) - \mu_1A_{ijk1}$$

$$\begin{aligned}\frac{dS_{ijk2}}{dt} &= -\lambda_{ijk2}S_{ijk2} + \gamma_{ij}Y_{ijk2} + \delta_{ij}Z_{ijk2} + \varphi_{ijk2}Z_{ijk2} \\ &\quad + p_{s_{ijk2}}\mu_1(S_{ijk1} + A_{ijk1}) - \mu_2S_{ijk2}\end{aligned}$$

$$\frac{dY_{ijk2}}{dt} = \sigma_{ij}\lambda_{ijk2}S_{ijk2} - \gamma_{ij}Y_{ijk2} - \mu_2Y_{ijk2} + \mu_1Y_{ijk1}$$

$$\frac{dZ_{ijk2}}{dt} = (1 - \sigma_{ij})\lambda_{ijk2}S_{ijk2} - \delta_{ij}Z_{ijk2} - \varphi_{ijk2}Z_{ijk2} - \mu_2Z_{ijk2} + \mu_1Z_{ijk1}$$

$$\frac{dA_{ijk2}}{dt} = (1 - p_{s_{ijk2}})\mu_1(S_{ijk1} + A_{ijk1}) - \mu_2A_{ijk2}$$

Cumulative reported cases (D) are calculated as:

$$\frac{dD_{ijkl}}{dt} = \pi_{s_{ij}}\gamma_{ij}Y_{ijkl} + \pi_a\varphi_{ijkl}Z_{ijkl}$$

Parameter values used in the model are provided in the main text (**Tables 1 and 2**).

Table 1: Parameter symbols and definitions.

Symbol	Definition
$i$	Subpopulation
$j$	Sex
$k$	Sexual activity group
$l$	Age group
$\mu$	Rate of model entry/exit
$\lambda$	Force of infection
$\sigma$	Probability symptomatic infection
$\gamma$	Recovery rate for symptomatic infection
$\delta$	Natural recovery rate for asymptomatic infection
$\varphi$	Screening rate for asymptomatic infection
$p_s$	Proportion ever sexually active
$\pi_s$	Probability symptomatic case is reported
$\pi_a$	Probability asymptomatic case is reported

### 3 Sexual mixing

Annual minimum rates of partner acquisition ( $c_{min}$ ) for a given age category and sex were estimated by model fitting, with estimates for MSM derived separately from other males. Prior estimates for relative rates of partner acquisition ( $rp$ ) in different population groups were estimated as described in section 5.1.1. For an individual of a given subpopulation ( $i$ ), sex ( $j$ ), sexual activity group ( $k$ ), and age group ( $l$ ), the mean annual rate of partner acquisition was calculated as:

$$c_{ijkl} = c_{min_{ijl}} r p_{ijkl}$$

The mixing matrix takes into account sexual activity group, age, and subpopulation membership. Within a given subpopulation ( $i$ ), we describe the probability that a person of sexual activity class  $k$  and age group  $l$  will form a partnership with a person of activity class  $s$  and age group  $t$  as:

$$\rho_{ijklst} = (\epsilon_{1,i} \delta_{ks} + (1 - \epsilon_{1,i}) \frac{\sum_{v=1}^2 N_{ij'sv} c_{ij'sv}}{\sum_{u=1}^2 \sum_{v=1}^2 N_{ij'uv} c_{ij'uv}}) * (\epsilon_{2,ijl} \delta_{lt} + (1 - \epsilon_{2,ijl}) \frac{\sum_{u=1}^2 N_{ij'ut} c_{ij'ut}}{\sum_{u=1}^2 \sum_{v=1}^2 N_{ij'uv} c_{ij'uv}})$$

where  $\delta_{xy} = 1$  if  $x = y$  and 0 otherwise.  $\epsilon_{1,i}$  defines mixing between sexual activity groups, while  $\epsilon_{2,ijl}$  defines mixing between age groups, with values ranging from 0 (random or proportionate mixing) to 1 (assortative or 'like with

like' mixing). Here,  $j'$  indicates that a partnership is formed with an individual of the opposite sex. For partnerships within the MSM population,  $j' = j$ .

A final parameter,  $\epsilon_{3,ij}$ , defines mixing between subpopulations; when equal to 1, all partnerships are formed with individuals belonging to the same subpopulation, and when equal to 0, all partnerships are formed with individuals from other subpopulations. For partnerships occurring outside of an individual's subpopulation, we assumed that the parameters describing age and sexual activity assortativity remained unchanged (i.e., preference for age or sexual risk group assortative mixing was the same, regardless of whether a partner was a member of the same subpopulation or not). Partnerships formed outside of an individual's own subpopulation were assumed to be distributed proportionally to the sizes of the other subpopulations. Here,  $m$  describes the subpopulation of the partner.

For  $i = m$ ,

$$\rho_{ijklmst} = \epsilon_{3,ij} \rho_{ijklst}$$

Otherwise (i.e., for  $i \neq m$ ),

$$\rho_{ijklmst} = (1 - \epsilon_{3,ij}) \frac{N_{mj'st}}{\sum_{w,w \neq i} N_{wj'st}} \rho_{ijklst}$$

Mixing between MSM and other subpopulations was assumed to occur via MSM forming sexual partnerships with females. We used the approach of Garnett and Anderson [10] to ensure that partnerships were balanced. When balancing partnership change rates within a subpopulation, we assumed that both sexes compromised equally, such that the number of partnerships formed was equal to the arithmetic mean of the number of partners desired by each sex, and used these adjusted partner change rates ( $c_{ijklmst}$ ) to calculate the force of infection (below).

To balance partnerships across subpopulations, we assumed that the group with the smaller population size determined the total number of partnerships formed. For example, the partner change rate between Hispanic females and non-Hispanic non-black males was determined by the desired number of partnerships in the females, since they represented a smaller proportion of the population.

## 4 Force of infection

The rate at which susceptible individuals are infected depends on the partner change rate ( $c_{ijklmst}$ ), the transmission probability per partnership ( $\beta_{ji}$ ) and the sexual mixing matrix ( $\rho_{ijklmst}$ ), where  $m$ ,  $s$ , and  $t$  represent the subpopulation, sexual activity group, and age category, respectively, of the sexual partner:

$$\lambda_{ijkl} = c_{rr} \beta_{ji} \sum_{m=1}^4 \sum_{s=1}^2 \sum_{t=1}^2 \frac{c_{ijklmst} \rho_{ijklmst} (Y_{mj'st} + Z_{mj'st})}{N_{mj'st}}$$

Note that  $j'$  indicates a partnership that is formed with an individual of the opposite sex. For partnerships formed within the MSM subpopulation, when  $m = i, j' = j$ . For heterosexual males, no partnerships are formed with MSM.  $c_{rr}$  is a time-varying term (estimated via model calibration) that represents relative risk of transmission in MSM, to account for changes in sexual risk behaviors over time; this parameter was assumed to be 1 in all other subpopulations.

## 5 Model calibration

We calibrated parameters describing sexual mixing, gonorrhea natural history, and screening rates using an adaptive Metropolis-Hastings Markov chain Monte Carlo algorithm implemented in R [11]. This method uses a Bayesian approach to estimate probability distributions for uncertain parameters, given the model and available data. This MCMC method allows for approximation of multi-dimensional distributions for which direct sampling is difficult or not possible. The adaptive procedure works to optimize the proposal distribution by first adapting the size of the covariance matrix to achieve an optimal acceptance rate, and then adapting the shape of the covariance matrix [12].

Prior parameter distributions were guided by the available data, using point estimates and plausible ranges obtained from the biomedical literature where possible, or based on expert opinion or assumption, as described in Tables 1 and 2. When information about parameters was scarce (e.g., sexual mixing coefficients), we assumed broad priors. We used the `riskDistributions` package [13] to estimate the parameters describing the prior probability distribution functions. Time-varying parameters (reporting probabilities and screening rates) were described by cubic Bézier curves, a type of parametric curve that allows for a sufficient degree of flexibility in the possible shapes it can take without requiring a large number of parameters to define it (described in more detail in 5.1.2). Parameters were either log transformed (to ensure positivity) or logit transformed (to ensure probabilities were bounded between 0 and 1). Additional details on model parameterization and calibration data targets are provided below.

The likelihood was specified as beta distributions around estimates of age assortativity, prevalence, reported case rates, proportion of male cases occurring in MSM, and proportion of cases reporting symptoms, and a normal distribution for relative reporting rates in black and Hispanic cases (data sources described in 5.2). Error variance was estimated from the data. For the diagnosed case data, we assumed 95% confidence intervals of  $\pm 20\%$  of the point estimates.

We initialized 10 independent MCMC chains of 100,000 iterations. These chains were visually assessed to ensure convergence and combined after burning and thinning. The prior and posterior distributions of the model parameters are presented in **Fig S2** and the results of model fitting are shown in **Fig 1** and **Fig S1**.

### 5.1 Parameters

#### 5.1.1 Sexual mixing parameters

The model included parameters describing assortativity of sexual mixing by age, sexual activity group, and subpopulation, which were used to calculate the

sexual mixing matrix [10]. For the heterosexual populations, all partnerships occurred with the opposite sex. For MSM, all within subpopulation mixing was with males, with bridging events occurring via sexual partnerships with females in the other subpopulations.

Estimates of prior values for relative rates of partner acquisition ( $rp$ ) were guided by 2011-2013 National Survey of Family Growth (NSFG) estimates of reported lifetime sexual partners [4]. For a given age category and racial/ethnic group, individuals in the 90<sup>th</sup> percentile of lifetime partners were assigned to the high sexual activity group, with the remainder of the population allocated to the low sexual activity group. We excluded individuals who reported no sexual partners in their lifetime. As the survey data are capped at a maximum of 50 lifetime partners, we fit censored log normal distributions to derive estimates of the median number of lifetime partners in the high and low activity groups. We used these values to estimate relative rates of partner change by sexual activity group and subpopulation for a given age and sex stratum.

### 5.1.2 Time-varying parameters

Screening rates and reporting probabilities were modeled as time-varying parameters, described by cubic Bézier curves. We did not model the screening and treatment processes separately for asymptomatic cases; the screening rate thus represents the probability that an asymptomatic case is detected and receives treatment. Each curve is defined by a start point ( $P_0$ ), end point ( $P_3$ ), and two internal control points ( $P_1$  and  $P_2$ ).

$$Y(t) = (1-t)^3P_0 + 3(1-t)^2tP_1 + 3(1-t)t^2P_2 + t^3P_3$$

We used a cubic Bézier interpolation to reparameterize the function to produce a smooth curve going through 4 equally spaced points ( $y_0, y_1, y_2, y_3$ ). The start and end points represent the values of screening/reporting at the beginning and end of the calibration period, respectively, while the two internal points determine the shape of the curve between the start and end points. Priors for the start and end points are provided in Table 2 of the main text. The two internal points were parameterized to fall between the start and end points:

$$\begin{aligned} y_1 &= y_0 + (y_3 - y_0)Beta(1, 1) \\ y_2 &= y_1 + (y_3 - y_1)Beta(1, 1) \end{aligned}$$

The Beta(1,1) distribution provided a broad prior, with the 95% interval spanning 0.025-0.975. These parameters are rescaled to control points for the Bézier function as:

$$\begin{aligned} P_0 &= y_0 \\ P_1 &= (-5y_0 + 18y_1 - 9y_2 + 2y_3)/6 \\ P_2 &= (2y_0 - 9y_1 + 18y_2 - 5y_3)/6 \\ P_3 &= y_3 \end{aligned}$$

These points are entered into the original cubic Bézier function to calculate the screening/reporting rate at each desired time point.

To limit the number of time-varying parameters estimated by model calibration, we assumed that the screening rates and reporting probabilities in some groups followed the same shape as those estimated as described above, but had

different magnitudes. This was implemented by multiplying these time-varying parameters by relative risks for specific population groups, as presented in **Table 2**. Specifically, reporting probabilities in symptomatic cases ( $\pi_{s_{ij}}$ ) were calculated as:

$$rr_{symp_{ijk}} * \pi_a$$

with  $rr_{symp_{ijk}}$  representing the relative risk a case is reported if symptomatic.

Screening rates ( $\varphi_{ijkl}$ ) were calculated as:

$$rr_{ack} * rr_{pop_{ij}} * \psi_{ijl}$$

where  $rr_{ack}$  is the relative risk of screening by sexual activity group,  $rr_{pop_{ij}}$  is the relative risk of screening by subpopulation and sex, and  $\psi_{ijl}$  is the time-varying estimate of annual screening rate for the low sexual activity group of a given age, sex, and subpopulation.

## 5.2 Data for model calibration

### 5.2.1 Sexual mixing data

We used estimates of mixing across age groups and subpopulations from the 2011-2013 NSFG [4]. The reported race/ethnicity of respondents' most recent sexual partner was used to estimate assortativity within subpopulations, with proportions stratified by sex. For sexually active survey respondents we obtained estimates of the proportion of most recent sexual partners that belonged to the same age category as the respondent for each age category and sex.

We did not observe significant differences across race/ethnicity in the proportion of respondents reporting age-assortative partners or differences across age groupings in the proportion of respondents reporting partners of the same race/ethnicity. As comparable measures were not available for MSM respondents, we used assumption to inform model priors describing sexual mixing in MSM.

### 5.2.2 Infection data

Data on gonorrhea prevalence and reported cases were used to ensure our model was reproducing observed trends. We used published data from the National Health and Nutrition Examination Survey (NHANES) (1999-2008) on age- and sex-specific gonorrhea prevalence [14]. Small sample sizes limit racial/ethnic prevalence comparisons, but estimates were available for non-Hispanic blacks [15]. We included a parameter (pNHANES) that was estimated by model fitting to allow for possible underrepresentation of MSM in the NHANES prevalence sample. This parameter represented the probability that a gonorrhea case occurring in MSM would be sampled for inclusion in prevalence calculations. For example, a value of pNHANES of 0.25 would indicate that only 1 in 4 prevalent cases in MSM would be counted in prevalence estimates.

Annual reported gonorrhea cases for the years 2000-2015 were obtained from the NCHHSTP Atlas [16]. Given changes in definitions of race/ethnicity over this time period, we used race/ethnicity-stratified data for the years 2011-2015 only. For each sex and age group, we calculated reported case rate ratios in non-Hispanic black and Hispanic populations, using overall rates as the referent.



National-level data on sex of sex partner are incomplete for gonorrhea cases. Available data from the STD Surveillance Network (SSuN) for the years 2010-2015 suggest that approximately 55% of male cases occur in MSM [*Mark Stenger, personal communication, note: no data available for 2014*]. We used these estimates of the proportion of MSM cases reported among males as an additional calibration target. Finally, we used data from SSuN on the proportion of diagnosed cases reporting symptoms as a measure of the expected distribution of detected cases by symptom status [17].

## 6 Model outputs

Since surveillance data for gonorrhea are based on reported cases, we modeled the treatment and reporting process in order have model outputs that were comparable to the available data. In the model, symptomatic cases are treated because they seek medical care for their symptoms. Asymptomatic cases are treated because they seek or are identified for screening. A treated case does not necessarily become a reported case, but all reported cases are assumed to have received treatment. For model fitting to surveillance data, we used model-generated reported cases. For estimation of the impact of alternate screening strategies, we used both incident and reported cases as the primary outcomes.

For each model simulation, infections averted (%) was calculated as:  $100 * (\text{total number of infections for the base case} - \text{total number of infections for the intervention}) / \text{total number infections for the base case}$ . Incidence was calculated as number of new gonorrhea infections per 100 population per year and reported case rates were calculated as number or reported gonorrhea infections per 100 population per year. Reported case rate ratios and incidence rate ratios by race/ethnicity were calculated as reported case rate or incidence rate for a given age-sex-subpopulation divided by overall population rates for a given age and sex. For comparisons of rates by race/ethnicity, we assumed cases in MSM were distributed across the three racial/ethnic groups, with cases allocated relative to the proportionate sizes of each of the racial/ethnic subpopulations. To account for the variability in the number of screening tests performed with the different strategies, and consequently the variable resource requirements, we calculated the number needed to screen to avert an incident gonorrhea infection (number of screening tests performed divided by the number of cases averted over the intervention time period).

## References

1. Centers for Disease Control and Prevention. Sexually transmitted disease surveillance 2015. Accessed 21 Feb 2017: <http://www.cdc.gov/std/stats> (2016).
2. National Center for Health Statistics. Vintage 2014 postcensal estimates of the resident population of the United States (April 1, 2010, Jul 1 2010–July 1, 2014), by year, county, single-year of age (0, 1, 2, ..., 85 years and over), bridged race, Hispanic origin, and sex. Accessed 27 Feb 2017. <[http://www.cdc.gov/nchs/nvss/bridged%7B%5C\\_%7Drace.htm](http://www.cdc.gov/nchs/nvss/bridged%7B%5C_%7Drace.htm)>.
3. Purcell, D. W. *et al.* Estimating the population size of men who have sex with men in the United States to obtain HIV and syphilis rates. *Open AIDS J* **6**, 98–107 (2012).
4. U.S. Department of Health and Human Services. Public Use Data File Documentation: 2011–2013 National Survey of Family Growth. Accessed 27 Feb 2017. <[https://www.cdc.gov/nchs/data/nsfg/nsfg%7B%5C\\_%7D2011-2013%7B%5C\\_%7Duserguide%7B%5C\\_%7Dmaintext.pdf](https://www.cdc.gov/nchs/data/nsfg/nsfg%7B%5C_%7D2011-2013%7B%5C_%7Duserguide%7B%5C_%7Dmaintext.pdf)>.
5. Elford, J. Changing patterns of sexual behaviour in the era of highly active antiretroviral therapy. *Current Opinion in Infectious Diseases* **19**, 26–32 (2006).
6. Hart, G. J. & Elford, J. Sexual risk behaviour of men who have sex with men: emerging patterns and new challenges. *Current Opinion in Infectious Diseases* **23**, 39–44 (2010).
7. Wiesner, P. J. & Thompson 3rd, S. E. Gonococcal diseases. *Dis Mon* **26**, 1–44 (1980).
8. McCormack, W. M., Stumacher, R. J., Johnson, K. & Donner, A. Clinical spectrum of gonococcal infection in women. *Lancet* **1**, 1182–1185 (1977).
9. Patton, M. E. *et al.* Extragenital gonorrhea and chlamydia testing and infection among men who have sex with men—STD Surveillance Network, United States, 2010–2012. *Clinical Infectious Diseases* **58**, 1564–1570 (2014).
10. Garnett, G. P. & Anderson, R. M. Balancing sexual partnerships in an age and activity stratified model of HIV transmission in heterosexual populations. *IMA J Math Appl Med Biol* **11**, 161–92 (1994).
11. Camacho, A. & Funk, S. fitR: Tool box for fitting dynamic infectious disease models to time series. R package version 0.1. <<http://sbfnk.github.io/mfiidd/fitR.tar.gz>> (2016).
12. Camacho, A. *et al.* Temporal changes in Ebola transmission in Sierra Leone and implications for control requirements: a real-time modelling study. *PLoS Curr* **7** (2015).
13. Belgorodski, N., Greiner, M., Tolksdorf, K. & Schueller, K. rriskDistributions: fitting distributions to given data or known quantiles. R package version 2.1.1. <<https://cran.r-project.org/package=rriskDistributions>> (2016).
14. Satterwhite, C. L. *et al.* Sexually transmitted infections among US women and men: prevalence and incidence estimates, 2008. *Sexually Transm Dis* **40**, 187 (2013).

15. Torrone, E. A. *et al.* Prevalence of *Neisseria gonorrhoeae* among persons 14 to 39 years of age, United States, 1999 to 2008. *Sexually Transm Dis* **40**, 202–205 (2013).
16. Centers for Disease Control and Prevention. NCHHSTP AtlasPlus. Accessed 27 Feb 2017. <<https://www.cdc.gov/nchhstp/atlas/>>.
17. Newman, L. M. *et al.* A tale of two gonorrhea epidemics: results from the STD surveillance network. *Public Health Rep* **127**, 282–292 (2012).